Architecture for ROVs detection and positioning based on hydrophone arrays

Alberto Belmonte-Hernández
Grupo de Aplicación de Telecomunicaciones visuales (GATV)
Universidad Politécnica de Madrid
Madrid, Spain
https://orcid.org/0000-0002-4009-2662

Anaida Fernández García
Grupo de Aplicación de Telecomunicaciones visuales (GATV)
Universidad Politécnica de Madrid
Madrid, Spain
https://orcid.org/0000-0002-6102-3121

César Antonio Ortiz-Toro
Escuela Técnica Superior de Ingenieros de Telecomunicación
Universidad Politécnica de Madrid
Madrid, Spain
https://orcid.org/0000-0002-7245-6328

Francisco Pérez Carrasco
FAV Innovation and Technologies Coop.V
Universitat Politècnica de València, Spain
https://orcid.org/0000-0001-9254-9962

Vicente Negro-Valdecantos
Escuela Técnica Superior de Ingenieros de Caminos, Canales y Puertos
Universidad Politécnica de Madrid
Madrid, Spain
https://orcid.org/0000-0002-5110-0891

Álvaro Gutiérrez Martín
Escuela Técnica Superior de Ingenieros de Telecomunicación
Universidad Politécnica de Madrid
Madrid, Spain
https://orcid.org/0000-0001-8926-5328

2025 6th International Conference in Electronic Engineering & Information Technology (EEITE) | 979-8-3315-4419-5/25/$31.00 ©2025 IEEE | DOI: 10.1109/EEITE65381.2025.11166125

***Abstract*— This paper presents an integrated architecture for detecting and positioning Remotely Operated Vehicles (ROVs) using strategically placed hydrophone arrays. We introduce two types of hydrophones with distinct sensitivity and bandwidth characteristics, deployed in both controlled (swimming pool) and semi-controlled (harbor) environments. The proposed system leverages frequency-domain (Discrete Fourier Transform, DFT) and time-frequency (Short-Time Fourier Transform, STFT) analyses to highlight key acoustic signatures, such as characteristic motor and propeller frequencies, enabling reliable ROV detection under variable noise conditions. We then incorporate propagation modeling and multi-hydrophone geometry to estimate the vehicle's position, drawing on both amplitude- and time-based techniques. The experimental results demonstrate robust performance in different aquatic settings, confirming the utility of high-sensitivity hydrophones and carefully selected signal processing parameters for enhanced detection and localization. Our findings emphasize how combining versatile hardware, advanced acoustic processing, and adaptive array configurations can address the challenges posed by complex underwater environments, laying the groundwork for broader applications such as harbor security, environmental monitoring, and maritime traffic management.**

Keywords—Hydrophones, System Architecture, Vessels Detection, Positioning, Acoustic processing.

I. INTRODUCTION

Underwater acoustic sensing has become an essential feature in a wide range of maritime applications, from environmental monitoring and marine mammal tracking to subsea navigation and autonomous underwater vehicle (AUV) localization. Central to these applications is the strategic placement and configuration of hydrophone arrays, which serve as underwater microphones tuned to pick up sound waves propagating through the water column. Depending on the sensing objective—be it passive listening for marine life, active acoustic imaging, or high-precision positioning—hydrophones can be arranged in various geometries that each exhibit unique advantages, drawbacks, and technical constraints.

A variety of baseline configurations exist to facilitate acoustic localization and navigation. For instance, Long Baseline (LBL) systems typically involve multiple fixed transmitters anchored in known locations, providing robust positioning over large regions but often requiring laborious deployment and calibration [1]. Short Baseline (SBL) and Ultra-Short Baseline (USBL) setups bring the hydrophones closer together, enabling more compact arrays that are simpler to handle yet may suffer from decreased positional accuracy or more complicated geometry when signals arrive at steep angles [2], [3], [4]. In harbor or shallow-water scenarios, arrays are often deployed in a planar configuration, whereas for deep-sea exploration, three-dimensional (3D) geometry with vertical spacing can improve depth estimation at the cost of higher deployment complexity [5], [6]. In each case, considerations such as the number of hydrophones, the available infrastructure, and the frequency range of interest play critical roles in determining the optimal layout.

Various real-world studies have explored hydrophone deployment strategies, balancing coverage, resolution, portability, and cost. For instance, a seabed-mounted network in a busy harbor enabled vessel and marine fauna monitoring but required frequent maintenance [7]. To address acoustic blind spots near the South Georgia Islands, a six-unit autonomous array was deployed in the Scotia Sea in 2008, enhancing signal detection in the Antarctic Convergence Zone [8]. Ocean Networks Canada (ONC) collaborates with coastal communities to deploy cabled and autonomous hydrophone systems, integrating icListen hydrophones in boat-based, buoy, and fixed observatories for long-term marine monitoring [9]. These varied approaches reflect trade-offs between coverage, resolution, portability, and cost in hydrophone placement strategies. Recent research on detecting vessels and propeller noise using hydrophones has employed a variety of signal processing and machine learning techniques. For example, spectral signature analysis and matched filtering have been used to identify the characteristic tonal frequencies of a rotating propeller in real-time monitoring systems [10]. Classification algorithms that exploit spectrograms as input to Deep Learning methods is an actual research solution which shows very promising results in several existing datasets [11], [12]. Additionally, for vessels positioning/tracking TDOA solutions and propagation model-based solutions have been widely adopted [13], [14].

This work explores the deployment of two hydrophone types in controlled (swimming pool) and semi-controlled (harbor) environments for underwater robot detection and localization. By analyzing frequency response, sensitivity, and placement geometry, we demonstrate how the chosen array effectively captures the robot's propulsion sound signatures, ensuring reliable position estimation even in challenging acoustic conditions.

II. PROPOSED ARCHITECTURE

A. Remotely Operated Vehicle (ROV) details

The Remotely Operated Vehicle (ROV) used in our test is the popular BlueROV2 [15] in its heavy configuration (Fig. 1). The BlueROV2 is a versatile, affordable, and powerful underwater vehicle designed for inspection and research missions. This compact and powerful ROV (457 x 338 x 254 mm, approx. 11-12 kg) is equipped with a robust frame, batteries, and high lumen lights for improved visibility of HD video feeds. This ROV uses four thrusters for horizontal locomotion and four thrusters for vertical locomotion, providing six degrees of freedom. The BlueROV2 supports operations to depths of 100 meters (extendable to 300 meters with modifications), achieves a maximum forward speed of 1.5m/s and can handle payloads of up to 1.4kg, depending on configuration.

This ROV is controlled by a Raspberry Pi and integrates an Inertial measurement unit (IMU), magnetometer and pressure sensor on an external expansion board. Using the open source ArduSub software, the BlueROV2 allows for autonomous capabilities and extensive customization, working with a variety of hardware including sonar sensors, cameras, and an inertial navigation system. While typically tethered, the open-source code allows for the development of a potential conversion to a ROV. In addition to the basic sensor configuration, each vehicle is equipped with an echo-sounder and mechanical scanning sonar for navigation and imaging. For the purposes of these tests, the control and management of the ROVs was performed using NAUSIM simulator [16].

Fig. 1. Photography of the BlueROV2 in its heavy configuration.

B. General data collection and processing architecture

Figure 2 shows the underwater acoustic detection system, which combines hardware and software to capture, process, and analyze underwater sounds. Hydrophones placed near maritime traffic continuously record acoustic signals, which are digitized and sent to Raspberry Pi 4b units. These devices segment the audio into 3-second WAV files for analysis. Data is transmitted efficiently via 5G or Ethernet, and results are displayed through a user-friendly control panel.

Step 1: The recorded audio files are stored in a cloud-based MinIO file storage system by the processing units. This storage solution allows the system to efficiently manage large amounts of audio data while ensuring quick access for analysis.

Step 2: The processing units send a message to the broker notifying that a file has been uploaded to the MinIO server. The message includes the timestamp, metadata about the audio file, and the location of the file in the server. The broker is deployed with RedPanda framework, and it manages communication between most components, enabling message reception and topic subscription through different protocols such as ampq, mqtt and Kafka compatibility.

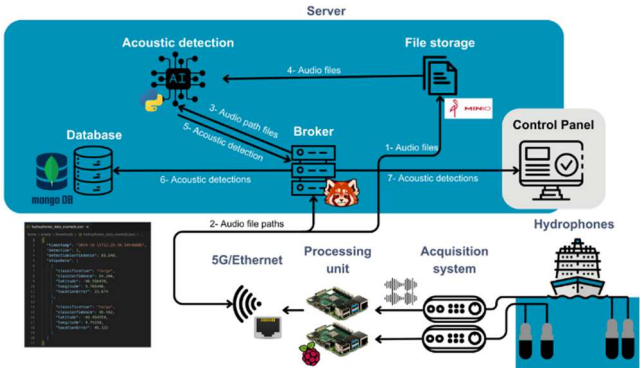

Fig. 2. Main system architecture and its components.

Step 3: The AI-powered acoustic detection module, implemented in Python, subscribes to a broker to retrieve audio files, then processes them to detect and classify vessel types, assign confidence scores, and estimate their relative position based on hydrophone data.

Step 4: The acoustic detection module sends the detection information to the broker. An example of the JSON structured file is shown in the bottom left corner of Figure 2.

Step 5: The database deployed with MongoDB stores all broker messages for future analysis and inquiry.

Step 6: The control panel receives the detection information or retrieve the historical one from the broker and the MongoDB to display the processed results in a user-friendly way.

The proposed architecture for acoustic detection is based on Figure 3, where a processing unit, an acquisition system and hydrophones are the basic components. Passive-acoustic sensors, i.e. hydrophones, gather information from the surroundings. This information is captured by a multi-input acquisition system, which provides a synchronized output that will be analyzed by a processing unit. One of the preliminary conditions for the system is that it must be modular and portable, being able to be placed and configured to different use case scenarios. Consequently, each component of the system must be lightweight, autonomous and robust.

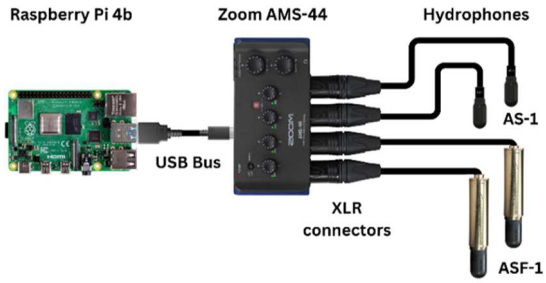

Fig. 3. Acoustic detection network system.

To capture underwater acoustic signals across a broad frequency and depth range, we used two hydrophone models with complementary features: the ASF-1 and the AS-1. Their detailed specifications are summarized in Table 1. While both cover similar frequency ranges (up to 100 kHz), the ASF-1 offers higher sensitivity (−192 dB re 1 µPa), making it better suited for detecting subtle mechanical signals without requiring an additional pre-amplifier. Its heavier, more robust building also provided greater stability in dynamic environments such as harbors. In contrast, the AS-1 is lighter

and rated for deeper operation (up to 200 m) but requires an external PA6 pre-amplifier with 48V phantom power. Given the context of our experiments—in swimming pools and near-shore waters—the ASF-1's superior sensitivity and mechanical robustness offered notable advantages.

Table 1. Hydrophones specifications comparison.

Specification	Aquarian AS-1	Ambient ASF-1 MKII
Frequency Response	1 Hz - 100 kHz ±2 dB	7Hz - 100 kHz ±2 dB
Sensitivity	-208 dB re 1V/µPa	-192dB re 1V/µPa
Polar pattern	Omnidirectional	Omnidirectional
Depth Rating	200 m	100 m
Power Requirement	Requires high-impedance amplifier	+48V phantom power
Housing material	Polyurethane encapsulant	Aluminum-brass
Weight	8g + 150g extra	270g

Moving into the data acquisition (DAQ) system, the audio interface Zoom AMS-44 was chosen providing USB data transfer for real-time processing, and 4 XLR/TRS inputs enabling phantom power through them, which is a 48V passive supply through the hydrophone's connectors pre-amplifiers. It also provides the option of operating autonomously without external power supply, just with batteries or connected by the USB data port to a computer, which is convenient for the port scenario where having external power supply is not assured. The integrated pre-amp equivalent input noise (EIN) is –121 dB and the resolution in bits per sample 24-bits/s, and maximum sampling frequency 94 kHz, which allows a maximum recorded frequency of 47 kHz. These specifications cover the needs of our application.

The processing unit has low computational requirements, as its role is mainly to collect sensory data and interface with the rest of the system. A Raspberry Pi 4b is used for this purpose due to its portability, efficiency, and versatility. Key specs include a Broadcom BCM2711 64-bit SoC @ 1.5 GHz, up to 8 GB RAM, wireless LAN, Gigabit Ethernet, 4 USB ports (2x3.0, 2x2.0), operating range of 0–50 ºC, and 5 V 3 A power input via USB-C. It is the only device requiring external power and also supplies power to the DAQ via USB. Audio is recorded using arecord with ALSA drivers, capturing multi-track wav files (24 bit, 94 kHz, 3 s) from each hydrophone channel. A Python service uploads each file to a MinIO server, notifies the broker with metadata, and deletes the local file to manage limited storage.

III. Data Collection and Processing

A. Analysis of recorded sounds by hydrophones

We are particularly interested in identifying distinct frequency components in the recorded data, such as those produced by the motors or propellers of the underwater robot. By examining both the frequency-domain representation (via the DFT) and the time-frequency representation (via the STFT, which produces spectrograms), we can highlight the dominant frequency bands where most of the robot's acoustic energy is concentrated. These analyses also help us to see transient behaviors, noise, and any other relevant frequency signatures.

The audio is sampled at $f_s = 94\ kHz$. For the DFT, we use segments of the signal in their entirety to reveal the overall frequency content. Meanwhile, for the Short-Time Fourier Transform (STFT), we choose an FFT size of $n_{fft} = 1024$ and a hop length of 512. These values strike a balance between frequency and time resolution. Specifically, with $n_{fft} = 1024$ at 94 kHz, the frequency resolution is approximately $\Delta f \approx \frac{94000}{1024} \approx 91.8Hz$, which is sufficiently fine to isolate dominant tonal components from the robot's motors. The hop length of 512 samples corresponds to a time step of $\frac{512}{94000} \approx 5.45$ ms. This partial overlap (50%) between consecutive frames captures transient events without sacrificing frequency discrimination. We also employ a logarithmic frequency axis in the spectrogram to better visualize both low-frequency and higher-frequency components that may arise from the robot's mechanical or propulsive elements. Overall, these parameter settings provide an appropriate compromise for time-frequency analysis at the given sampling rate.

Figure 4 represents the results of a recording lasting approximately 12 minutes, during which the underwater robot performed multiple trajectories and remained stationary for extended periods. These recordings were captured using the hydrophones ASF-1, and the AS-1. The ASF-1, with its broader frequency response and higher sensitivity captures finer details of the acoustic signal, while the AS-1, although less sensitive, still provides adequate performance for identifying key acoustic features. The DFT analysis reveals distinct frequency components generated by the robot's propulsion system. The ASF-1 shows clear peaks at multiple frequencies, particularly in the low and mid-frequency ranges, where motor and propeller harmonics dominate. These peaks are sharper and of higher magnitude compared to the AS-1, highlighting the ASF-1's superior sensitivity. Nonetheless, the AS-1 also captures the fundamental frequencies and harmonics, confirming its utility for detecting key features of the robot's acoustic signature. The spectrograms provide a time-frequency view of the recordings, showing both the temporal variations and frequency content. The ASF-1 captures more detailed structures, including distinct horizontal bands corresponding to motor and propeller noise, and periods of reduced spectral energy during the robot's extended stops. The AS-1, while less detailed, still captures the main acoustic features, including the same extended stops and movement phases.

Key frequencies observed in the analysis include low frequencies (<1000 Hz), likely related to the primary motor operations, and mid frequencies (1000–10,000 Hz), where propeller harmonics are most prominent. These frequencies are critical for tracking the robot's movement and identifying its operating states. Higher frequencies (>10,000 Hz) are better captured by the ASF-1 and provide finer details about mechanical noise or turbulence, though these components are less pronounced in the AS-1 recordings. To enhance detection and tracking, the audio was segmented into 3-second intervals. This duration provides a balance between capturing key acoustic features, such as motor harmonics and transient events, while enabling efficient processing for real-time applications. Short segments allow the system to adapt quickly to changes in the robot's movement or acoustic environment, while still preserving sufficient frequency and time resolution. These intervals are particularly suitable for identifying periodic signals and optimizing the performance of detection and localization algorithms. [11], [12].

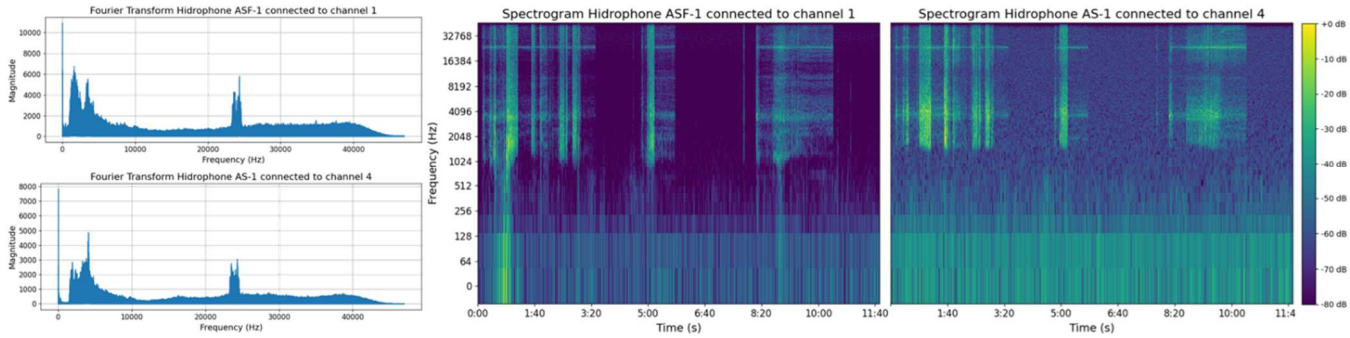

Fig. 4. Analysis with FFT and with Logarithmic Mel Spectrograms of ASF-1 and AS-1 Hydrophones in a 12-minute recording.

Figure 5 represents a 3-second audio segment spectrogram recorded by two hydrophones: the ASF-1 (Channel 0, left) and the AS-1 (Channel 3, right). Both spectrograms clearly capture distinct tonal components in the low-to-mid frequency range, particularly between 500 Hz and 10 kHz, which are indicative of the robot's motor and propeller harmonics. The ASF-1 demonstrates superior sensitivity, with higher energy levels across all frequency bands, making subtle spectral details more evident, particularly in the higher frequency ranges above 10 kHz. In contrast, the AS-1 spectrogram, while slightly noisier and less detailed in the upper frequencies, still reliably detects the dominant tonal features and broadband noise, showing its capability for detection and tracking tasks. The results highlight the importance of the ASF-1 for detailed acoustic analysis, while the AS-1 provides a cost-effective alternative with sufficient performance for detecting the primary acoustic features of interest. This analysis supports the use of short 3-second segments, as they effectively capture transient and continuous acoustic signatures critical for positioning and tracking.

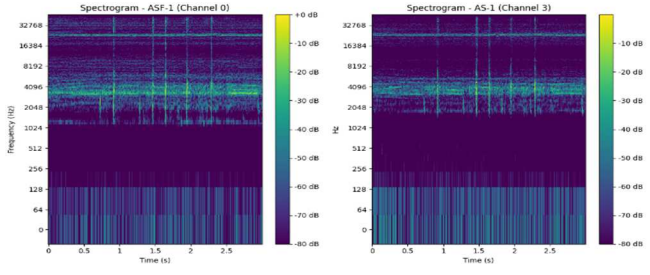

Fig. 5. 3 seconds LogMel spectrograms with ASF-1 and AS-1.

B. Data processing for detection and positioning

The positioning algorithm developed for the swimming pool environment relies on the estimation of the received signal amplitude at multiple hydrophones and the calibration of a propagation model. The first step involves selecting specific frequency bands associated with the underwater robot's propulsion system. These bands were identified through spectral analysis, ensuring they represent consistent acoustic emissions from the ROV rather than ambient noise. For each hydrophone, the signal amplitude A_i is extracted from these frequency bands using a STFT, isolating the energy within the predefined range. Once the amplitude is obtained, the distance d_i between the robot and each hydrophone is estimated using a propagation model. The acoustic signal's attenuation underwater is modeled by incorporating geometric spreading and absorption losses:

$$A_i = A_0 d_i^{-n} e^{-\alpha d_i}$$

where A_0 is the reference amplitude at a known distance, n is the geometric spreading factor (empirically adjusted based on experimental data), and α represents the absorption coefficient of the medium. Given that the swimming pool is a confined environment with reflective boundaries, the parameters n and α were adjusted based on control measurements at known distances, ensuring an optimal fit to the experimental data. Specifically, a set of calibration points at known distances between the ROV and the hydrophones was used to collect amplitude data. A nonlinear least squares fitting process was then applied to these data points to minimize the error between the measured and predicted amplitudes, yielding optimal estimates for n and α.

With the estimated distances d_i from each hydrophone, the position of the robot is computed using a nonlinear least squares trilateration approach. The objective is to minimize the error function:

$$min \sum_{i=1}^{N} \left(\sqrt{(x - x_i)^2 + (y - y_i)^2} - d_i \right)^2$$

where (x, y) is the estimated position of the ROV, and (x_i, y_i) are the known positions of the hydrophones. The optimization was performed using a least-squares solver, which iteratively adjusts (x, y) to minimize the discrepancy between the modeled and observed distances. To evaluate the accuracy of the positioning algorithm, we conducted controlled experiments by placing the ROV at predefined coordinates within the pool. The estimated positions were then compared to the ground-truth locations to calculate mean and maximum distance errors, validating the precision of the method.

For the detection algorithm, we build upon previous work that has explored vessel classification using spectrogram-based analysis under different noise conditions [17]. This approach leverages convolutional neural networks (CNNs) for feature extraction and classification, demonstrating high accuracy in distinguishing between various vessel types. As future work, we aim to extend this methodology to a multi-class detection framework, incorporating additional vessel categories, including the ROV used in our experiments. Furthermore, we plan to enhance the system by integrating source separation techniques, allowing the differentiation of overlapping acoustic signals from multiple sources. This extension will improve the robustness of the detection algorithm, making it suitable for complex real-world environments where multiple vessels and background noise are present. For the scope of this paper and its presentation, the presence of the ROV will be identified by detecting the previously analyzed frequency bands with a high acoustic signal power, using an empirically determined threshold.

IV. RESULTS

A. Deployment in water tank

The first tests were conducted within the test tank of the Ports Laboratory, in the Civil Engineering Technical School premises of UPM, a three-dimensional testing pool that has a depth of 1.36 meters, 11 meters wide, and 33 meters long. The bottom is finished with in situ terrazzo flooring, and the vertical walls are covered in continuous tiles, equipped with an anti-reflection system using 10 ppi polyurethane foam. Multidirectional waves are generated with absorption control, with individual blade movement, capable of filling up to 1.00 meter, but they were not used for our initial tests.

Fig. 5. Photographs of the water tank for the experiments.

For the test, we used a section of the tank that was 17 meters long and 11 meters wide. The deployment of the hydrophones involved 2 ASF-1 and 4 AS-1 hydrophones, 2 audio interface and 2 Raspberries to cover the whole area and use the complete set. The set of 4 AS-1 hydrophones were placed covering the 4 furthest corners of the area of the tank in use. The other 2 hydrophones were located at 8.5 m of the longest side of the tank. All of them were separated 25 cm from the closer wall to avoid undesired noises due to hits against the wall and were hanging 1 m from the top of the pool. The area was divided into 6 subareas as shown in Fig 6, to annotate the position of the ROV as reference for future analysis. The study performed consisted in the ROV doing different trajectories along the tank, getting closer and away from each hydrophone, while its subarea was annotated manually. In this way, we managed to record 6 different audio tracks, one per hydrophone, completely covering 187 m^2 where the ROV is moving around. This configuration is an ideal study case to characterize the sound produced by the ROV without other interferences and test the initial positioning algorithms.

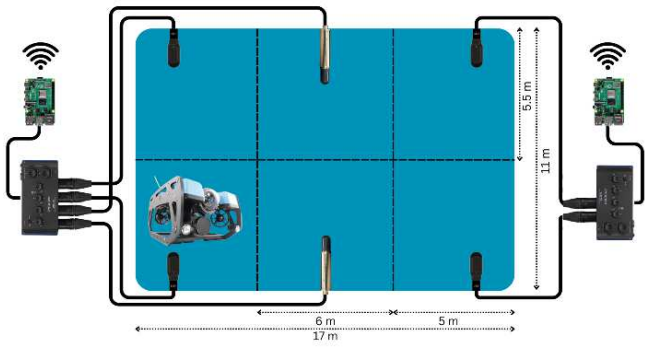

Fig. 6. Deployment of the system in the water tank and quadrant-based division for ROV position estimation.

The model parameters were adjusted separately for each hydrophone type to account for their differences in sensitivity and frequency response. For the ASF-1, which has higher sensitivity, the propagation model was calibrated using an initial reference amplitude A_0 at 1 meter, a spreading factor of $n = 1.8$, and an absorption coefficient of $\alpha = 0.002\ Np/m$, ensuring a good fit with measured signal attenuation. For the AS-1, due to its lower sensitivity, a higher empirical threshold was required, leading to $n = 2.2$ and $\alpha = 0.004\ Np/m$ to better match the observed decay in signal power. These values were derived by analyzing the spectrograms and FFT results, ensuring that the estimated distances were consistent with the known positions of the hydrophones during calibration experiments.

The accuracy of quadrant classification was evaluated based on manual annotations, resulting in a 79% accuracy. This value is not entirely representative, as the annotations were approximate and transitions between quadrants introduce significant uncertainty. When the robot moves across quadrant boundaries, misclassification is likely due to the inherent imprecision of manual labeling. Nonetheless, the use of a redundant system with six hydrophones, instead of the minimum required three for trilateration, helps mitigate these errors by providing additional constraints and improving the robustness of position estimation. To contextualize the performance of our method, we compared it against a baseline approach using simple geometric trilateration with constant sound speed and no attenuation model achieving 47% accuracy in quadrant classification. Our method demonstrated superior accuracy, particularly in the reflective pool environment, where signal attenuation and reverberation significantly affect simpler models.

B. Deployment in Valencia Port

The tests in Valencia were conducted in a 60 × 70-meter area at the north arm of the Xitá dock (Fig. 7). Deploying in a real port presented challenges such as power availability, weather conditions, interferences, and access restrictions. To address these, a single deployment point was established with one AS-1 and one ASF-1 hydrophone connected to an acquisition system housed in an IP65 enclosure on the dock, powered by a portable battery. Two external cameras were included to capture visual context alongside the acoustic recordings. The hydrophones were mounted on a PVC structure to keep them separated from the dock wall, minimizing unwanted reflections and interferences.

Fig. 7. Deployment of the system in Valencia Port and equipment prepared to work outdoors.

In this test, we also relied on the maneuvers of the ROV within the area of 60 x 70 m, in which the operations that could be performed entailed larger distances that the test in the water tank, therefore the limitations of the acquisition system could be better analyzed, together with the interferences of the port activities and the comparison between AS-1 and ASF-1 hydrophones in a realistic environment. On the other hand, as we only had one placement for both hydrophones, the positioning estimation could not be addressed with the proposed approach.

The recorded power spectral density (PSD) of the harbor and pool environments shows distinct background noise profiles. Figure 8 presents the mean PSD from 10 samples (3 seconds each) taken during periods of no visible activity, covering the 1 Hz–4 kHz range. In the pool, ambient noise decreases from −75 dB to −105 dB between 0–1 kHz, then remains stable. In contrast, the harbor displays a more complex profile, fluctuating between −85 and −95 dB, with a sharp drop beyond 3.25 kHz. This complexity is likely due to vessel activity and the open, acoustically reflective environment. These conditions make detection and localization more difficult.

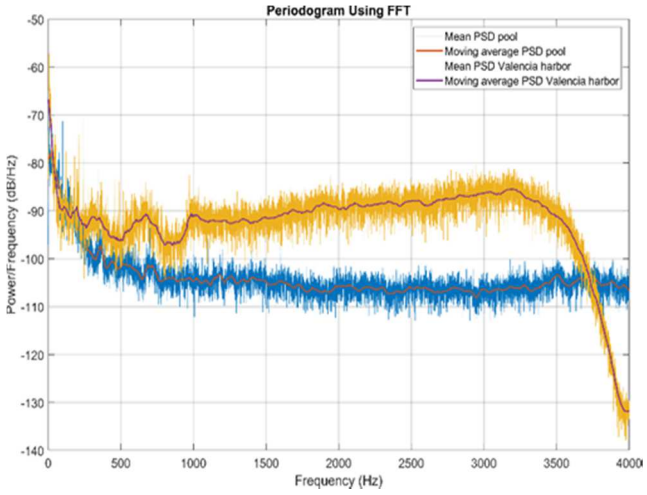

Fig. 8. Power spectral density of ambient noise for both the pool and Valencia environments.

In response, the next phase of this work will focus on implementing noise reduction techniques and source separation methods to isolate the ROV's acoustic signature from interfering signals. This enhancement is essential for adapting the system to real-world conditions, where uncontrolled noise sources can impact detection accuracy.

V. Conclusions and Future Work

This study demonstrates the feasibility of using hydrophone arrays for the detection and positioning of ROVs in controlled and semi-controlled environments. By combining frequency and time-frequency analysis with amplitude-based propagation modeling, we achieved reliable tracking of the ROV's movements, even in the presence of acoustic reflections and background noise. The experimental results highlight the advantages of using high-sensitivity hydrophones for improved signal clarity and the benefits of a redundant hydrophone setup in mitigating localization errors.

Future work will focus on expanding the detection system to a multi-class classification framework, incorporating additional vessel types beyond the ROV and, we are analyzing techniques such as diffusion model for noise cleaning in more challenging scenarios. This will involve refining deep learning models for vessel recognition and integrating source separation techniques to improve detection in environments with overlapping acoustic sources. Further experiments will also be conducted in larger and more dynamic maritime settings to validate the scalability and robustness of the proposed approach.

Acknowledgment

This work has been supported by the European Commission within the context of the project SMAUG, funded under EU Horizon Europe Grant Agreement 101121129. This work has partially been supported by Grant PID2020-112502RB-C41 and grant PID2023-146540OB-C42 funded by MCIN/AEI/10.13039/501100011033.

References

[1] Y. Zhu and L. Zhou, "Hybrid Tightly-Coupled SINS/LBL for Underwater Navigation System," in IEEE Access, vol. 12, pp. 31279-31286, 2024, doi: 10.1109/ACCESS.2021.3051398.

[2] Yuyi, Zhai et al. "Study of Underwater Positioning Based on Short Baseline Sonar System." 2009 International Conference on Artificial Intelligence and Computational Intelligence 2 (2009): 343-346.

[3] Reis, J., Morgado, M., Batista, P., Oliveira, P., & Silvestre, C. (2016). Design and Experimental Validation of a USBL Underwater Acoustic Positioning System. Sensors, 16(9), 1491.

[4] Bo Xu, Haibin Zhu, Yu Guo, A robust iterative algorithm for SINS/USBL integrated navigation based on dual hydrophone differential model, Measurement, Volume 242, Part A, 2025, 115854, ISSN 0263-2241.

[5] S. Zhu, Y. Yang, Y. Wang, Q. Yang and X. Yan, "Performance of a cylindrical hydrophone array for practical use," 2017 IEEE. International Conference on Signal Processing, Communications and Computing (ICSPCC), Xiamen, China, 2017, pp. 1-5.

[6] Dai, Y., & Sun, C. (2023). Adaptive Beamforming with Hydrophone Arrays Based on Oblique Projection in the Presence of the Steering Vector Mismatch. Journal of Marine Science and Engineering, 11(4), 876.

[7] Dewey, Richard & Dakin, Tom & Mouy, Xavier & Urazghildiiev, Ildar. (2015). A REGIONAL HYDROPHONE NETWORK: MONITOR, DETECT AND TRACK.

[8] Matsumoto, Haru & Bohnenstiehl, D. & Dziak, Bob & Embley, Robert & Park, Minkyu. (2008). DEPLOYMENT OF AUTONOMOUS HYDROPHONE ARRAY IN THE SCOTIA SEA Cooperative Institute for Marine Resources Studies Sponsored by National Nuclear Security Administration.

[9] R. Flagg, J. Kanes, L. Marshall, K. Holman, D. Bulger and J. Duggan, "Ocean Acoustic Data Collection - Community-Based Methods for Meeting Community Needs," OCEANS 2022, Hampton Roads, Hampton Roads, VA, USA, 2022, pp. 1-6

[10] Zhu, C., Seri, S. G., Mohebbi-Kalkhoran, H., & Ratilal, P. (2020). Long-Range Automatic Detection, Acoustic Signature Characterization and Bearing-Time Estimation of Multiple Ships with Coherent Hydrophone Array.

[11] L. C. F. Domingos, P. E. Santos, P. S. M. Skelton, R. S. A. Brinkworth and K. Sammut, "An Investigation of Preprocessing Filters and Deep Learning Methods for Vessel Type Classification With Underwater Acoustic Data," in IEEE Access, vol. 10, pp. 117582-117596, 2022.

[12] David Santos-Domínguez, Soledad Torres-Guijarro, Antonio Cardenal-López, Antonio Pena-Gimenez, ShipsEar: An underwater vessel noise database, Applied Acoustics, Volume 113, 2016, Pages 64-69, ISSN 0003-682X.

[13] Tesei, Alessandra et al. "Tracking of multiple surface vessels based on passive acoustic underwater arrays." The Journal of the Acoustical Society of America 147 2 (2020).

[14] MacGillivray, Alexander O. and Christ A. F. de Jong. "A Reference Spectrum Model for Estimating Source Levels of Marine Shipping Based on Automated Identification System Data." Journal of Marine Science and Engineering (2021).

[15] B. Robotics "BlueROV2: The world's most affordable high-performance ROV. BlueROV2" Datasheet; Blue Robotics: Torrance, CA, USA. (2016).

[16] C. A. Ortiz-Toro et al., "Exploring UUV Development with NauSim: An Open-Source Simulation Platform," 2024 Global Conference on Wireless and Optical Technologies (GCWOT), Malaga, Spain, 2024, pp. 1-7, doi: 10.1109/GCWOT63882.2024.10805602.

[17] Hjelmervik, KT et al. "DATA AUGMENTATION AND PEPROCESSING TECHNIQUES FOR ENHANCED UNDERWATER DETECTION AND CLASSIFICATION." International Conference on Underwater Acoustics 2024.